JOURNAL OF RESEARCH of the National Bureau of Standards
Volume 82, No. 2, September–October 1977

Minimum Polynomials and Control in Linear Systems*

John Z. Hearon**

(June 13, 1977)

Given the constant coefficient system $\dot{x} = Ax + Bu$, relationships are established among the minimum polynomial (with respect to A) of the range of B, the degree and null space of this polynomial, the rank of the controllability matrix and the degree of the minimum polynomial of A. These relations lead to a simple proof of a theorem on reduction of control.

Key words: Control theory; linear algebra; matrix; minimum polynomials.

1. Introduction

Given a set of vectors $\mathbf{B} = \{b_1, b_2, \ldots, b_r\}$ and a square matrix A, we consider the following objects: The minimum polynomial $\phi(A)$ of the linear span of $\mathbf{B}$; the dimension, ρ, of the A-invariant subspace generated by $\mathbf{B}$; and the dimension, ν, of the null space of $\phi(A)$. Three lemmas are given which relate the degree of ϕ, the degree of the minimum polynomial of A, and the quantities ρ and ν.

The motivation and natural context for the lemmas is linear control theory. The lemmas are used to prove a theorem in control theory. The result is known but the proof here is very simple as well as elementary (which is not the same thing).

Both the concept and language of controllability matrices and controllable subspaces have recently appeared frequently [1, 10, 12, 13][1] in the context of abstract linear algebra. Thus the lemmas of this paper are of possible interest in their own right.

2. Preliminaries

In what follows the underlying vector space is the set of all complex n-tuples. All matrices considered have complex entries. For any matrix M we denote by $R(M)$, $\rho(M)$ and $N(M)$ the range, rank (dimension of the range) and null space respectively. For any column vector $\mathbf{y}$ we denote by $\mathbf{y}^*$ the conjugate transpose of $\mathbf{y}$. Employing classical terminology, we call a square matrix nonderogatory whenever the minimum polynomial of A coincides with the characteristic polynomial. Such a matrix is sometimes called cyclic [7]. A nonderogatory matrix can be characterized by its elementary divisor structure or by its Jordan form (e.g., see [11] p. 13). It is easy to see that a matrix is nonderogatory if and only if the characteristic polynomials of its Jordan block are pairwise relatively prime. Equivalently, a matrix is nonderogatory if and only if distinct Jordan blocks involve distinct roots of the characteristic polynomial. In the literature of control theory this latter property is often invoked (e.g., see [8], p. 86) and the term nonderogatory not used at all.

Given a square matrix A and a vector $\mathbf{v}$, the minimum polynomial of $\mathbf{v}$ with respect to A is that monic polynomial, p, of minimal degree such that $p(A)\mathbf{v} = 0$. When the matrix A is understood or clear from context we simply refer to the minimum polynomial of a vector. The minimum polynomial of $\mathbf{v}$ is clearly unique and it divides without remainder any polynomial which annihilates $\mathbf{v}$ [2, 6, 7]. In principle the minimum polynomial of $\mathbf{v}$ is constructed as follows: consider the sequence defined by $\mathbf{v}_0 = \mathbf{v}$, $\mathbf{v}_i = A\mathbf{v}_{i-1}^2 = A^i\mathbf{v}$, $i = 1, 2, 3, \ldots$. There will be a first vector which is linearly dependent on the preceding ones. Let it be $\mathbf{v}_s$. Then for some scalars $\alpha_0, \alpha_1, \ldots, \alpha_s$, not all zero with $\alpha_s = 1$, we have $\alpha_0\mathbf{v}_0 + \alpha_1\mathbf{v}_1 + \ldots + \alpha_{s-1}\mathbf{v}_{s-1} + \mathbf{v}_s = p(A)\,\mathbf{v} = 0$, where $p(\lambda) = \alpha_0 + \alpha_1\lambda + \ldots + \alpha_{s-1}\lambda^{s-1} + \lambda^s$. Thus $p(A)$ annihilates $\mathbf{v}$ and by construction is the monic polynomial of least degree which does so.

Given a subspace S, the minimum polynomial of S (with respect to a matrix A) is the monic polynomial, w, of least degree such that $w(A)\mathbf{x} = 0$, for each $\mathbf{x} \epsilon \mathbf{S}$. If we determine the minimum polynomial for each vector of a basis set in S, then clearly the least common multiple of these polynomials is the minimum polynomial

* An invited paper. "Manuscript originally received September 1975," (Revised March 5, 1976).
** Present address: Mathematical Research Branch, NIAMDD, National Institutes of Health, Bethesda, Md. 20014.

[1] Figures in brackets indicate the literature references at the end of the paper.

of S. The minimum polynomial of a subspace S, since it is an annihilating polynomial for each vector in S, contains as a divisor the minimum polynomial of every vector in S.

It is a standard theorem [2, 6, 7, 9] that given an n-square matrix A, there exists in n-space a vector whose minimum polynomial coincides with the minimum polynomial of A (which, of course, is the minimum polynomial of the entire space). Some of these proofs [2] [6], but not others [7] [9], are easily adapted to prove that in any proper subspace there exists a vector whose minimum polynomial coincides with the minimum polynomial of the subspace. This is a result which we will need in what follows. Recently [5] an elegant proof has been given but the proof pivots on a quite technical lemma. The proof (for the whole space) in [6] is, characteristically, extremely short and it draws only on the concepts discussed in this section. For completeness, we give here the (trivial) modification which adapts this proof to a (possibly) proper subspace and we render the logical sequence somewhat less succinct.

THEOREM 1: *Let* S *be any subspace and* A *a given square matrix. Then there exists a vector* $\mathbf{v} \epsilon$ S *such that the minimum polynomial of* v *is the minimum polynomial of* S.

PROOF (Householder): We assume the dimension of S to be two or more, since otherwise the theorem is obvious. Let $\mathbf{v} \epsilon \mathbf{S}$ be a vector whose minimum polynomial, f, is of maximal degree in S. Let $\mathbf{u} \epsilon \mathbf{S}$ be any vector, linearly independent of $\mathbf{v}$, with minimum polynomial g. Then w, the least common multiple of f and g, annihilates every vector in the (u, v)-plane and thus contains as divisor the minimum polynomial of every vector in the (u, v)-plane. But w has finitely many divisors and thus there must exist in the (u, v)-plane independent vectors $\mathbf{a}$ and $\mathbf{b}$ each with the same minimum polynomial, h. Thus, h annihilates every vector in the (a, b)-plane which is the (u, v)-plane. In particular h annihilates v. As an annihilating polynomial of $\mathbf{v}$, h is divisible by f and hence degree $h \geq$ degree f. But as the minimum polynomial of a vector in S, degree $h \leq$ degree f, since the degree of f is maximal. Thus $f = h$. It follows that f annihilates every vector in the (u, v)-plane, and since u was arbitrary (so long as it was not a multiple of v) f annihilates S. Plainly no polynomial of lesser degree can do so, and f is the minimum polynomial of S.

A linear, constant coefficient system

$$\dot{x} = Ax + Bu, \tag{1}$$

where A is n-square and B is nxr, is said to be (completely) controllable if for each pair of vectors $\mathbf{a}$, $\mathbf{b}$ there exists a control vector $\mathbf{u}$ such that $\mathbf{x} = \mathbf{a}$ at $t = 0$ and $\mathbf{x} = \mathbf{b}$ at $t = T$, for some finite T. The system (1) is controllable if and only if the matrix

$$C = [B, AB, A^2B, \cdots, A^{n-1}B] \tag{2}$$

has rank n. The matrix C is defined as the controllability matrix and its range $R(C)$ as the controllability space. The controllability space is obviously the smallest subspace, invariant under A, containing $R(B)$. For, we observe that any A-invariant subspace containing $R(B)$ contains each column of C. When the system (1) is controllable (not controllable) it is sometimes said that (A,B) is controllable (not controllable) and we use this terminology.

3. Lemmas and Theorem

Consider an n-square matrix A and an nxr matrix B. Let ϕ be the minimum polynomial, with respect to A, of the subspace $R(B)$ and the degree of ϕ be d. We denote by $N(\phi)$ the null space of $\phi(A)$ and denote the dimension of $N(\phi)$ by ν. We denote the degree of the minimum polynomial of A by m. The matrix C is defined as in (2).

LEMMA 1:

$$\nu \geq \rho(\mathrm{C}) \geq \mathrm{d} \leq \mathrm{m}.$$

PROOF: Let $\mathbf{x} \epsilon \mathbf{N}(\phi)$. Then $\phi(A)A\mathbf{x} = A\ \phi(A)\mathbf{x} = 0$ and hence $\mathbf{Ax} \epsilon N(\phi)$. Thus $N(\phi)$ is an A-invariant subspace and it clearly contains $R(B)$ which $\phi(A)$ annihilates. But $R(C)$ is the smallest such subspace and we have $R(C) \subseteq N(\phi)$ and hence $\nu \geq \rho(C)$. By Theorem 1, there is at least one vector, $\mathbf{b}$, in $R(B)$ whose minimum polynomial coincides with ϕ. Let this vector be augmented to a basis for $R(B)$ and let $\hat{B}$ be the

matrix whose columns are these basis vectors. If $\hat{C} = [\hat{B}, A\hat{B}, \cdots, A^{n-1}\hat{B}]$, then $\rho(C) = \rho(\hat{C})$, since the columns of C and of $\hat{C}$ clearly span the same subspace. But $\hat{C}$ has at least d linearly independent columns, namely $\mathbf{b}, A\mathbf{b}, \cdots, A^{d-1}\mathbf{b}$, since the minimum polynomial of b is of degree d. Thus we have $\rho(C) = \rho(\hat{C}) \geq d$. Finally, it is trivial that d never exceeds m and this completes the proof.

LEMMA 2: *We have* $\nu = \mathrm{n}$ *if and only if* $\mathrm{d} = \mathrm{m}$.

PROOF: By Lemma 1 and the standard rank-nullity Theorem we have

$$\rho(C) \leq n - \rho(\phi) = \nu \tag{3}$$

where $\rho(\phi)$ is the rank of $\phi(A)$. From this inequality, $\nu = n$ implies $\rho(\phi) = 0$ or that $\phi(A) = 0$. But this requires $d = m$. For, $\phi(A) = 0$ means that ϕ is divisible by Ψ the minimum polynomial of A while the degree, d, of ϕ never exceeds m. Conversely assume $d = m$. We observe that Ψ is always divisible by ϕ. For, Ψ annihilates all of $R(B)$, is thus divisible by the minimum polynomial of every vector in $R(B)$ but, by Theorem 1, there is a vector in $R(B)$ whose minimum polynomial is ϕ. Thus when ϕ and Ψ are of the same degree we must have $\phi = \Psi$, which means $\nu = n$. This completes the proof.

LEMMA 3: *If* (A,B) *is controllable, then*

$$\mathrm{n} = \nu = \rho(\mathrm{C}) \geq \mathrm{d} = \mathrm{m}.$$

PROOF: If (A,B) is controllable then $\rho(C) = n$. Since $\nu \leq n$, always, we read from Lemma 1 that $\nu = n$. This being the case, $d = m$ follows from Lemma 2. This completes the proof.

REMARK: From (3), we can have a very direct proof, without Lemma 2, that $\rho(C) = n$ implies $d = m$. For, assume $d < m$. Then, $\rho(\phi)$ cannot be zero and (3) tells that $\rho(C) \leq n-1$.

Given that the system (1) is controllable with a vector controller, u, it *may* be possible to control the system by a choice such as $\mathbf{u} = \mathbf{Fy}$, where F is an rxk matrix, $k < r$, and $\mathbf{y}$ is a $\mathbf{k}$-vector. This reduction of the control space from r to k dimensions, when possible, is of obvious advantage both practically and theoretically. In particular, since many control problems depend solely upon the property of controllability, theoretical analysis may be simplified by dealing with the reduced system.

The ultimate in control reduction is the case $k = 1$, when $F = \mathbf{c}$ is an $\mathbf{r}$-vector and $\mathbf{y}$ is a scalar controller: $\mathbf{u(t) = c}y(t)$. In this case we inquire for a vector $\mathbf{b} \epsilon \mathbf{R}(B)$ such that the system

$$\dot{x} = Ax + by \tag{1'}$$

is controllable whenever the system (1) is controllable. Necessary and sufficient conditions for this follow readily from the lemmas and are set forth in the following theorem.

THEOREM 2: *The following conditions are equivalent.*

(i) *The pair* (A, B) *is controllable and* A *is nonderogatory*
(ii) $\mathrm{d} = \mathrm{n}$
(iii) *There exists a* $\mathrm{b} \epsilon \mathrm{R(B)}$ *such that* (A, b) *is controllable.*

PROOF: That (i) implies (ii) follows from Lemma 3 by setting $m = n$. That (ii) implies (i) follows from Lemma 1 by setting $d = n$. If we assume (ii), then by Theorem 1 there is a vector $\mathbf{b} \epsilon \mathbf{R(B)}$ whose minimum polynomial is of degree $d = n$. It follows that (A, b) is controllable and that (ii) implies (iii). If there exists a vector $\mathbf{b} \epsilon \mathbf{R(B)}$ such that (A, b) is controllable then that vector clearly cannot have minimum polynomial of degree less than n. Thus no polynomial of degree less than n can annihilate $R(B)$. Thus $d \geq n$ and hence $d = n$, so that (iii) implies (ii). This completes the proof of the theorem.

In [8], theorem 6, p. 86, it is shown that given the controllability of (1) then (1′) is controllable for some $\mathbf{b} \epsilon \mathbf{R(B)}$ if and only if A is nonderogatory. (Observe that we have shown that controllability of (1′) for some $\mathbf{b} \epsilon \mathbf{R(B)}$ implies both that A is nonderogatory and that (1) is controllable) The proof in [8] is very involved and hinges on extensive Jordan form manipulation. The implications (i) $\Leftrightarrow$ (iii) can be deduced from a more general but fairly complicated theorem due to Heymann [4, theorem 2, p. 565]. The implication (i) $\Rightarrow$ (iii) follows from theorem 2 of [3].

If we are interested only in Theorem 2, the several quite direct proofs are available. We give one such alternative proof: That (i) implies (ii) follows at once from the Remark. That (ii) implies (iii) is proved as in the proof of Theorem 2. Given (iii), we have $n = m$. For, otherwise $b, Ab, \ldots, A^{n-1}b$ would be linearly dependent as expressed by $\Psi(A)b = 0$. We also have $\rho(C) = n$, since the linear span of $b, Ab, \ldots, A^{n-1}b$ is contained in $R(C)$. Thus (iii) implies (i).

4. References

[1] Carlson, D., and Loewy, R., On ranges of Lyapunov transformations, Lin. Alg. and Appl. **8,** 237–48 (1974).
[2] Gantmacher, F. R., The Theory of Matrices, Vol. I (Chelsea, New York, 1959).
[3] Hautus, M. L. J., Controllability and observability conditions for linear autonomous systems, Nederl. Akad. Wet. Proc. **A72,** 443–48 (1969).
[4] Heymann, M., On input and output reducibility of multivariable linear systems, IEEE Trans. Aut. Control. AC **15,** 563–69 (1970).
[5] Heymann, M., and Thorpe, J. A., A note on minimum polynomials, Lin. Alg. and Appl. **7,** 279–80 (1973).
[6] Householder, A. S., The Theory of Matrices in Numerical Analysis, (Blaisdell, New York, 1964).
[7] Jacobson, N., Lectures in Abstract Algebra, Vol. II, (D. Van Nostrand, New York, 1953).
[8] Lee, E. B., and Markus, L., Foundations of Optimal Control Theory, (J. Wiley, New York, 1967).
[9] Paige, L. J., and Swift, J. D., Elements of Linear Algebra, (Ginn, 1961).
[10] Snyders, J., and Zakai, M., On nonnegative solutions of the equation $AD + DA' = -C^*$, SIAM J. Appl. Math **18,** 704–14 (1970).
[11] Wilkinson, J. H., The Algebraic Eigenvalue Problem, (Clarendon, Oxford, 1965).
[12] Wimmer, H. K., Inertia theorems for matrices, controllability and linear vibrations, Lin. Alg. and Appl. **8,** 337–43 (1974).
[13] Wimmer, H. K., An inertia theorem for tridiagonal matrices and a criterion of Wall on continued fractions, Lin. Alg. and Appl. **9,** 41–4 (1974).